COMMUNICATING WITH TEAMS OF COOPERATIVE ROBOTS

D. Perzanowski, A.C. Schultz, W. Adams, M. Bugajska, E. Marsh, G. Trafton, and D. Brock
Codes 5512, 5513, and 5515, Naval Research Laboratory, Washington, DC 20375

M. Skubic
University of Missouri-Columbia, Computer Engineering & Computer Science Department, Columbia, MO 65211

M. Abramson
ITT Industries, Alexandria, VA 22303

Abstract: We are designing and implementing a multi-modal interface to a team of dynamically autonomous robots. For this interface, we have elected to use natural language and gesture. Gestures can be either natural gestures perceived by a vision system installed on the robot, or they can be made by using a stylus on a Personal Digital Assistant. In this paper we describe the integrated modes of input and one of the theoretical constructs that we use to facilitate cooperation and collaboration among members of a team of robots. An integrated context and dialog processing component that incorporates knowledge of spatial relations enables cooperative activity between the multiple agents, both human and robotic.

Keywords: cooperative and collaborative behaviour, dynamic autonomy, human-robot interaction, multi-modal interfaces

1. INTRODUCTION

Interacting and communicating with another person is a complicated set of processes in real life. However, humans learn and master the linguistic and social skills necessary to perform this feat with seemingly little effort.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2002		**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Communicating with Teams of Cooperative Robots	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,4555 Overlook Ave SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT
We are designing and implementing a multi-modal interface to a team of dynamically autonomous robots. For this interface, we have elected to use natural language and gesture. Gestures can be either natural gestures perceived by a vision system installed on the robot, or they can be made by using a stylus on a Personal Digital Assistant. In this paper we describe the integrated modes of input and one of the theoretical constructs that we use to facilitate cooperation and collaboration among members of a team of robots. An integrated context and dialog processing component that incorporates knowledge of spatial relations enables cooperative activity between the multiple agents, both human and robotic.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**9**	

In just a few short years they are capable of carrying on conversations and other interactions with another human and in most cases have little difficulty extending these skills to perform similar functions with a group of individuals. Certain aspects of communication enable individuals to form teams to achieve their goals. We are interested in investigating what those aspects of communication are, and then incorporate them in our multi-modal interface for human-robot interactions.

We have already incorporated natural language and gestures into our human-robot interface [Perzanowski et al., 1998; 2000]; we are now introducing additional context and dialog processing to facilitate natural communication and enable cooperative action between multiple agents.

2. COMMUNICATION ISSUES

2.1 Linguistic Cues

Certain contextual and linguistic cues provide crucial information to humans for them to communicate easily. Prosodic cues, such a the inflection of one's voice, the rise and lowering of pitch of the voice, tell the participants of a dialog that an utterance is being made, a certain type of utterance is being made, and that the utterance is ending, or more is to come. However, state-of-the-art speech recognition engines sensitive to this kind of information are not commercially available. It would seem that greater cooperation and teamwork between humans and robots is stymied by the inability of speech recognition engines to provide important information to participants in a dialog. However, other cues used by humans enable them to interact and exchange information during a dialog. We currently use the syntactic and semantic information that both our speech recognition system, ViaVoice, and natural language understanding system, Nautilus (Wauchope, 1994), provide. Additional contextual information is obtained from visual cues, spatial information and an analysis of the linguistic information available to us in *context predicate*s (Perzanowski et al., 1999) to foster collaboration and cooperation in a team of human and robot agents. We turn now to a discussion of these features.

2.2 Visual Cues

Visual cues, such as "body language," provide humans with the kinds of information needed to facilitate dialog and promote teamwork. For example, if the speaker of sentence (1) is standing in front of two individuals but

staring at one of them, then it is incumbent upon the person being stared at to respond in some way.

(1) The computer is over there.

Likewise, the speaker of (1) might gesture—point—or simply shrug a shoulder in a particular direction to indicate information about the location of the object.

Finally, participants in a dialog may either directly address whom they wish to perform certain actions, as in (2), or they may focus their attention on a person or a thing.

(2) Coyote, go to the computer on the left side of the room.

Eye gaze directed at Coyote, without directly addressing Coyote in (2), cues all the listeners of the utterance to the fact that the speaker wishes Coyote to perform the action. Nodding one's head at the listener can indicate the same intentions. Therefore, visual cues can be utilized by an interface to compensate for the lack of certain information.

2.3 Knowledge

Knowledge of the various participants and the environment can also facilitate collaborative communication. For example, if someone knows that a person can only make group meetings on Fridays at 10 o'clock, a great deal of extraneous communication can be avoided, given such a pre-condition. Likewise, knowledge of the capabilities—the strengths and/or weaknesses—of the various agents in a dialog can benefit communication.

Asking someone to lift an object when that person is not capable of doing so is counter-productive. Likewise, if one of the sensors on a robot team member suddenly fails and is no longer usable, sharing this information with the other participants can prevent extraneous communication and wasting time.

Environmental information, such as spatial knowledge (Skubic, et al. 2002), can also assist team members in achieving their goals. Determining that an object is within range of the sensors of one robot, and having that robot communicate this information to the other participants, contributes to a more timely solution to the task.

In our initial research, we focused on natural language and natural gestures in command and control situations with a single robot or multiple robots that still acted independently. We are working with a team of dynamically autonomous robots interacting without constant human intervention. We define the term "dynamically autonomous" to mean that the agents are capable of operating at varying levels of autonomy, based on their individual awareness of their own capabilities in achieving some goal; their awareness of other agents' capabilities; and their knowledge of the

overall plan (Pollack and McCarthy, 1999) and the history of achieving subgoals as the overall plan progresses (Grosz, et al. 1999).

3. MULTI-MODAL INTERFACE

3.1 Gesture and Object Recognition

We are using several robots, Nomad 200s, XR-4000s, and an RWI ATRV-Jr. Gestures are detected using a structured light rangefinder. A camera fitted with a filter tuned to the laser wavelength is mounted on its side. The robot is capable of tracking the user's hands and interpreting their motion as vectors or measured distances. A more detailed discussion can be found in (Perzanowski et al. 1998). Sonar sensors on the robots detect objects in the environment. With this data, object recognition is possible (Skubic et al 2001a; 2001b). We are currently incorporating a binocular vision system to permit more sophisticated recognition of both objects and people.

The interface (Figure 1) also employs a PDA with a stylus and touch-screen. Pointing, clicking or drawing on the touch screen indicate locations, regions, directions, and the like.

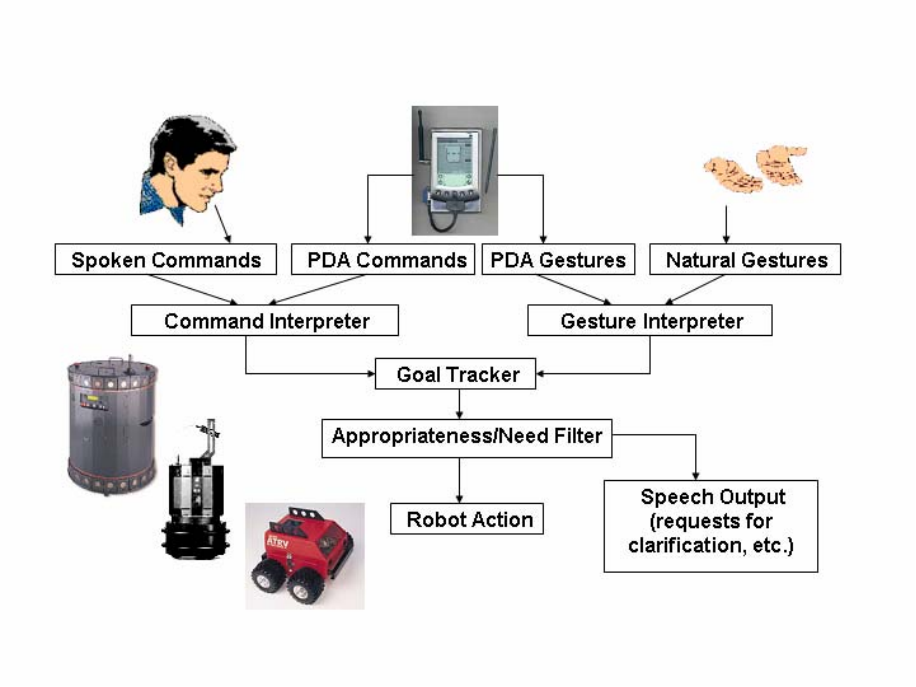

Figure 1. Multi-modal Interface

3.2 Natural Language Processing

A more detailed description of the natural language processing is discussed elsewhere (Perzanowski et al. 1998), but a brief discussion here introduces one element of the dialog which we employ for collaboration and cooperation in achieving goals. Vocal commands or clicks on buttons on the PDA screen are mapped into a logical form. The latter is correlated with gesture data, knowledge of the other participating agents, and with spatial information obtained from the robot sensors. The result is then mapped to a robot command, which produces either some action or an interchange of information. For example, the human user can direct a robot by uttering sentence (3).

(3) Coyote, go to the north side of the nearest building.

The spatial reasoning component uses the sensor data to determine that an object exists and it computes where the north side of the object is. If the sensors detect an appropriate object, the various inputs are combined and a robot command is sent to the robot to act accordingly. If, on the other hand, no such object is sensed, the robot complains verbally, saying something to the effect that no such object exists.

We track information about goals, i.e. whether or not goals have been attained, in *context predicates*. Context predicates are linguistically motivated constructs that contain semantic and contextual information of the discourse. (4) is the context predicate for (3).

```
(4) ((imper (:verb gesture-go
            (:agent (:system you))
            (:to-loc ((:thing side)
                      (:dir north))
                     ((:relation-to building)
                      ((:descrip nearest)
                       (:relation-to you))))))  1))
```

If a goal is achieved, the context predicate reflects this, as signified by the "1" in the representation. If the goal is not achieved, the representation exhibits a "0." As the discourse continues, the stack of context predicates is updated: if the focus of the dialog changes, completed goals are eliminated, but non-completed goals remain. Since this knowledge is shared by all of the participants in the dialog, anyone can act upon the non-completed goals, if the situation warrants it. Thus, if for some reason Coyote is unable to complete its specified goals, another robot can be tasked to complete the goals.

As the dialog progresses, the focus of the dialog changes (Grosz and Sidner, 1986). Keeping track and updating the focus of the dialog updates the context predicates.

We are currently interested in having robots determine on their own--based upon a particular task, their individual capabilities, knowledge and overall plan (Grosz, et al. 1999)--what teams should be formed, and who is a member of which team. Tasks can be achieved with as little human intervention as possible. Once the initial task is given, robots can form their own groups and obtain the goals more easily because they group themselves according to their individual strengths and appropriateness for completing certain goals. Thus, for example, armed robots would determine that they would be the best candidates for certain kinds of operations, while robots not so equipped would be more appropriate candidates for other missions. Furthermore, if one robot is tasked to go to a building, but another is closer, we are building in the capability to permit the latter robot to intervene and perform the action.

4. RELATED WORK

We are attempting to incorporate linguistic and visual information into a multi-media interface to foster collaborative and cooperative teamwork.

Other models incorporating collaboration and discourse theory exist, such as COLLAGEN (Rich, et al. 2001) and TRIPS (Allen, et al. 2001). Like COLLAGEN, we are grounding our work in linguistic and discourse theory and attempting to make the interface application-independent. However, we incorporate context predicates from the discourse, and unlike COLLAGEN we are using visual cues and spatial information to motivate team formation and teamwork.

TRIPS already incorporates much of the collaborative kinds of interaction we are looking for in a dialog. However, with our emphasis on context predicates, we are hoping to minimize human intervention in the collaboration.

Our emphasis on multi-modal and natural interaction sets us somewhat apart from the work of (Fong, et al. 2001). This research does not emphasize natural language in their interface to control a robot, and natural gestures are not employed. Instead, their interactions are limited to a set of messages and their gesturing is viewed as a translation of gestures into a visual joystick. We, on the other hand, are interested in natural commands and visual interactions with robotic agents. While our work incorporating a PDA device is very similar, we have not attempted any interface with a Web-based interface at this time. However, our goal is identical: development of a system in which humans and robots work together as cooperative agents in performing some task.

5. FUTURE WORK

While we do not incorporate a Web-based interface presently, we are working on adding this capability. In the future, we hope to access online information about novel locations, so that the robots can navigate through unknown terrain, having obtained information about routes and the environment from internet sources.

We are currently expanding our knowledge component to incorporate vocabulary acquisition in real-time. At present, if an object is sensed, and the human user tells a robot that the object is called a "computer," for example, the spatial reasoning component maintains this information, but it is not passed to the natural language understanding component. In other words, while the object "computer" exists in a robot's sensor readings and in its knowledge of the space around it, it still cannot communicate information about the computer naturally. Simply, while it knows that a computer exists, it cannot talk about it, or perform some rather rudimentary reasoning about the object so labelled.

We are, therefore, working on adding the ability to reason about objects. Thus, if an object is perceived from a certain viewpoint, we are adding the ability to know that an object, let's say a computer, is the same computer if viewed from a different point of view. We would also like for our team of robots to know that objects once identified, if moved, are still the same objects. Only their locations have changed.

We continue to focus our attention on the use of context predicates and a dialog-based planning component to motivate team formation and teamwork

6. CONCLUSION

We are concentrating on two main research areas to facilitate cooperation and collaboration in a team of robots. The first area focuses on context predicates, linguistically motivated constructs that contain semantic and goal information. Using context predicates, teams of robots share information about goal status and act accordingly. The second research area is our expansion of the spatial reasoning component so that robots reason about their physical environment and share information about the environment, objects, and locations.

Our purpose is to enhance team formation and dynamic autonomy so that robots interact with each other and human intervention occurs only as needed.

ACKNOWLEDGMENTS

The Naval Research Laboratory and the Office of Naval Research partly funded this research.

REFERENCES

Allen, J., Byron, D.K., Dzikovska, M., Ferguson, G., Galescu, L., and Stent, A. 2001. Toward Conversational Human-Computer Interaction. *AI Magazine,* (22)4:27-37.

Fong, T., Thorpe, C., and Baur, C. 2001. Advance Interfaces for Vehicle Teleoperation: Collaborative Control, Sensor Fusion Displays, and Remote Driving Tools. *Autonomous Robots,* 11: 77-85.

Grosz, B. and Sidner, C. 1986. Attention, Intentions, and the Structure of Discourse. *Computational Linguistic*s, 12(3):175–204.

Grosz, B., Hunsberger, L. and Kraus, S. 1999. Planning and Acting Together. *AI Magazin*e, 20(4): 23–34.

Perzanowski, D., Schultz, A.C. and Adams, W. 1998. Integrating Natural Language and Gesture in a Robotics Domain. In *Proc. IEEE Int'l Symp. Intelligent Control*, Piscataway, NJ, pp. 247–252.

Perzanowski, D., Schultz, A., Adams, W., and Marsh, E. 1999. Goal Tracking in a Natural Language Interface: Towards Achieving Adjustable Autonomy. In *Proc. 1999 IEEE Int'l Symp. Computational Intelligence in Robotics and Automation*, Piscataway, NJ, pp. 144–149.

Perzanowski, D., Adams, W., Schultz, A., and Marsh, E. 2000. Towards Seamless Integration in a Multimodal Interface. In *Proc. 2000 Workshop Interactive Robotics and Entertainmen*t, Menlo Park, CA, pp. 3–9.

Pollack, M. and McCarthy, C. 1999. Towards Focused Plan Monitoring: A Technique and an Application to Mobile Robots. In *Proc. 1999 IEEE Int'l Symp. Computational Intelligence in Robotics and Automatio*n, Piscataway, NJ, pp. 144–149.

Skubic, M., Perzanowski, D., Schultz, A., and Adams, W. 2002. Using Spatial Language in a Human-Robot Dialog. In *2002 IEEE Int'l Conf. on Robotics and Automation.*

Skubic, M., Chronis, G., Matsakis, P., and Keller, J. 2001a. Generating Linguistic Spatial Descriptions from Sonar Readings Using the Histogram of Forces. In *Proc. of the 2001 IEEE Int'l Conf. on Robotics and Automation*, Seoul, Korea.

Skubic, M., Chronis, G., Matsakis, P., and Keller, J. 2001b. Spatial Relations for Tactical Robot Navigation. In *Proc. of the SPIE, Unmanned Ground Vehicle Technology III*, Orlando, FL.

Rich, C., Sidner, C., and Lesh, N. 2001. COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction. *AI Magazine,* 22(4):15-25.

Wauchope, K. 1994. *Eucalyptus: Integrating Natural Language Input with a Graphical User Interfac*e, Technical Report NRL/FR/5510-94-9711, Naval Research Laboratory, Washington, D.C.